On the regularization of singular c-optimal designs

Luc Pronzato

On the regularization of singular c-optimal designs*

Luc Pronzato

Laboratoire I3S, CNRS/Université de Nice Sophia-Antipolis
Bât. Euclide, Les Algorithmes
BP 121, 2000 route des lucioles
06903 Sophia Antipolis cedex, France

November 26, 2008

Abstract

We consider the design of c-optimal experiments for the estimation of a scalar function $h(\theta)$ of the parameters θ in a nonlinear regression model. A c-optimal design ξ^* may be singular, and we derive conditions ensuring the asymptotic normality of the Least-Squares estimator of $h(\theta)$ for a singular design over a finite space. As illustrated by an example, the singular designs for which asymptotic normality holds typically depend on the unknown true value of θ, which makes singular c-optimal designs of no practical use in nonlinear situations. Some simple alternatives are then suggested for constructing nonsingular designs that approach a c-optimal design under some conditions.

Keywords. singular design, optimum design, c-optimality, D-optimality, regular asymptotic normality, consistency, LS estimation

AMS Subject Classification. 62K05, 62E20

1 Introduction

We consider experimental design for least-squares estimation in a nonlinear regression model with scalar observations

$$Y_i = Y(x_i) = \eta(x_i, \bar{\theta}) + \varepsilon_i\,, \quad \text{where } \bar{\theta} \in \Theta\,,\ i = 1, 2 \ldots \tag{1}$$

where $\{\varepsilon_i\}$ is a (second-order) stationary sequence of independent random variables with zero mean,

$$\mathbb{E}\{\varepsilon_i\} = 0 \text{ and } \mathbb{E}\{\varepsilon_i^2\} = \sigma^2 < \infty \ \ \forall i\,, \tag{2}$$

*This paper is dedicated to Andrej Pázman on the occasion of his 70th birthday

Θ is a compact subset of $\mathbb{R}^p$ and $x_i \in \mathcal{X}$ denotes the design point characterizing the experimental conditions for the i-th observation Y_i, with $\mathcal{X}$ a compact subset of $\mathbb{R}^d$. For the observations $Y_1, \ldots, Y_N$ performed at the design points $x_1, \ldots, x_N$, the Least-Squares Estimator (LSE) $\hat{\theta}^N_{LS}$ is obtained by minimizing

$$S_N(\theta) = \sum_{i=1}^{N} [Y_i - \eta(x_i, \theta)]^2 \,, \tag{3}$$

with respect to $\theta \in \Theta \subset \mathbb{R}^p$. We suppose throughout the paper that either the x_i's are non-random constants or they are generated independently of the Y_j's (i.e., the design is not sequential). We shall also use the following assumptions:

$\mathbf{H1}_\eta$: $\eta(x, \theta)$ is continuous on Θ for any $x \in \mathcal{X}$;

$\mathbf{H2}_\eta$: $\bar{\theta} \in \mathrm{int}(\Theta)$ and $\eta(x, \theta)$ is two times continuously differentiable with respect to $\theta \in \mathrm{int}(\Theta)$ for any $x \in \mathcal{X}$.

Then, under $\mathbf{H1}_\eta$ the LS estimator is strongly consistent, $\hat{\theta}^N_{LS} \stackrel{\text{a.s.}}{\to} \bar{\theta}$, $N \to \infty$, provided that the sequence $\{x_i\}$ is "rich enough", see, e.g., [3]. For instance, when the design points form an i.i.d. sequence generated with the probability measure ξ (which is called a randomized design with measure ξ in [7, 9]), strong consistency holds under the estimability condition

$$\int_{\mathcal{X}} [\eta(x, \theta) - \eta(x, \bar{\theta})]^2 \, \xi(dx) = 0 \ \Rightarrow \theta = \bar{\theta} \,. \tag{4}$$

Under the additional assumption $\mathbf{H2}_\eta$, $\hat{\theta}^N_{LS}$ is asymptotically normally distributed,

$$\sqrt{N}(\hat{\theta}^N_{LS} - \bar{\theta}) \stackrel{\text{d}}{\to} z \sim \mathcal{N}(\mathbf{0}, \mathbf{M}^{-1}(\xi, \bar{\theta})) \,, \ N \to \infty \,, \tag{5}$$

provided that the information matrix (normalized, per observation)

$$\mathbf{M}(\xi, \bar{\theta}) = \frac{1}{\sigma^2} \int_{\mathcal{X}} \left. \frac{\partial \eta(x, \theta)}{\partial \theta} \right|_{\bar{\theta}} \left. \frac{\partial \eta(x, \theta)}{\partial \theta^\top} \right|_{\bar{\theta}} \xi(dx) \tag{6}$$

is nonsingular.

The paper concerns the situation where one is interested in the estimation of $h(\theta)$ rather than in the estimation of θ, with $h(\cdot)$ a continuous scalar function on Θ. Then, when the estimability condition (4) takes the relaxed form

$$\int_{\mathcal{X}} [\eta(x, \theta) - \eta(x, \bar{\theta})]^2 \, \xi(dx) = 0 \ \Rightarrow h(\theta) = h(\bar{\theta}) \,, \tag{7}$$

we have $h(\hat{\theta}^N_{LS}) \stackrel{\text{a.s.}}{\to} h(\bar{\theta})$, $N \to \infty$. Under the assumption

$\mathbf{H}_h$: $h(\theta)$ is two times continuously differentiable with respect to $\theta \in \mathrm{int}(\Theta)$,

assuming, moreover, that $\partial h(\theta)/\partial \theta|_{\bar{\theta}} \neq \mathbf{0}$ and that (5) is satisfied, we also obtain (see [5, p. 61])

$$\sqrt{N}[h(\hat{\theta}^N_{LS}) - h(\bar{\theta})] \stackrel{\text{d}}{\to} \omega \sim \mathcal{N}\left(0, \left. \frac{\partial h(\theta)}{\partial \theta^\top} \right|_{\bar{\theta}} \mathbf{M}^{-1}(\xi, \bar{\theta}) \left. \frac{\partial h(\theta)}{\partial \theta} \right|_{\bar{\theta}} \right) \,, \ N \to \infty \,. \tag{8}$$

In Sect. 2 we prove a similar result on the asymptotic normality of $h(\hat{\theta}_{LS}^N)$ when $\mathbf{M}(\xi,\bar{\theta})$ is singular, that is,

$$\sqrt{N}[h(\hat{\theta}_{LS}^N) - h(\bar{\theta})] \xrightarrow{\text{d}} \omega \sim \mathcal{N}\left(0, \left.\frac{\partial h(\theta)}{\partial \theta^\top}\right|_{\bar{\theta}} \mathbf{M}^-(\xi,\bar{\theta}) \left.\frac{\partial h(\theta)}{\partial \theta}\right|_{\bar{\theta}}\right), \; N \to \infty\,, \quad (9)$$

with $\mathbf{M}^-$ a g-inverse of $\mathbf{M}$. This is called *regular asymptotic normality* in [9], where it is shown to hold under rather restrictive assumptions on $h(\cdot)$ but without requiring $\hat{\theta}_{LS}^N$ to be consistent. We show in Sect. 2 that when the design space $\mathcal{X}$ is finite $\hat{\theta}_{LS}^N$ is consistent under fairly general conditions, from which (9) then easily follows.

We use the standard approach and consider an experimental design that minimizes the asymptotic variance of $h(\hat{\theta}_{LS}^N)$. According to (9), this corresponds to minimizing $[\partial h(\theta)/\partial \theta^\top\big|_{\bar{\theta}}]\, \mathbf{M}^-(\xi,\bar{\theta})\, [\partial h(\theta)/\partial\theta\big|_{\bar{\theta}}]$. Since $\bar{\theta}$ is unknown, local c-optimal design is based on a nominal parameter value θ^0 and minimizes $\phi_c(\xi) = \Phi_c[\mathbf{M}(\xi,\theta^0)]$ with

$$\Phi_c(\cdot) : \mathbf{M} \in \mathbb{M}^{\geq} \to \begin{cases} \mathbf{c}_{\theta^0}^\top \mathbf{M}^- \mathbf{c}_{\theta^0} & \text{if and only if } \mathbf{c}_{\theta^0} \in \mathcal{M}(\mathbf{M}) \\ \infty & \text{otherwise} \end{cases} \quad (10)$$

where $\mathbb{M}^{\geq}$ denotes the set of non-negative definite $p \times p$ matrices,

$$\mathcal{M}(\mathbf{M}) = \{\mathbf{c} : \exists \mathbf{u} \in \mathbb{R}^p\,,\ \mathbf{c} = \mathbf{M}\mathbf{u}\}$$

and

$$\mathbf{c}_{\theta^0} = \left.\frac{\partial h(\theta)}{\partial \theta}\right|_{\theta^0}.$$

Note that the value of $\Phi_c(\mathbf{M})$ is independent of the choice of the g-inverse $\mathbf{M}^-$. Nonlinearity may be present in two places, since the model response $\eta(x,\theta)$ and the function of interest $h(\theta)$ may be nonlinear in θ. Local c-optimal design corresponds to c-optimal design in the linear (or more precisely linearized) model $\eta_L(x,\theta) = \mathbf{f}_{\theta^0}^\top(x)\theta$ where $\mathbf{f}_{\theta^0}(x) = \partial\eta(x,\theta)/\partial\theta\big|_{\theta^0}$, with the linear (linearized) function of interest $h_L(\theta) = \mathbf{c}_{\theta^0}^\top\theta$. A design ξ^* minimizing $\phi_c(\xi)$ may be singular, in the sense that the matrix $\mathbf{M}(\xi^*,\theta^0)$ is singular. In spite of an apparent simplicity for linear models, this yields, however, a difficulty due to the fact that the function $\Phi_c(\cdot)$ is only lower semi-continuous at a singular matrix $\mathbf{M} \in \mathbb{M}^{\geq}$. Indeed, this property implies that

$$\lim_{N\to\infty} \mathbf{c}^\top \mathbf{M}^-(\xi_N)\mathbf{c} \geq \mathbf{c}^\top \mathbf{M}^-(\xi)\mathbf{c}$$

when the empirical measure ξ_N of the design points converges weakly to ξ, see e.g. [6, p. 67] and [8] for examples with strict inequality. The two types of nonlinearities mentioned above cause additional difficulties in the presence of a singular design: both $\hat{\theta}_{LS}^N$ and $h(\hat{\theta}_{LS}^N)$ may not be consistent, or the asymptotic normality (9) may not hold, see [8] for an example with a linear model and a nonlinear function $h(\cdot)$. It is the purpose of the paper to expose some of those difficulties and to make suggestions for regularizing a singular c-optimal design.

2 Asymptotic properties of LSE with finite $\mathcal{X}$

When using a sequence of design points i.i.d. with the measure ξ, the condition (4) implies that $S_N(\theta)$ given by (3) grows to infinity at rate N when $\theta \neq \bar{\theta}$ (an assumption used in the classic reference [3]). On the other hand, for a design sequence with associated empirical measure converging to a discrete measure ξ, this amounts to ignoring the information provided by design points $x \in \mathcal{X}$ with a relative frequency $r_N(x)/N$ tending to zero, which therefore do not appear in the support of ξ. In order to acknowledge the information carried by such points, we can follow the same approach as in [10] from which we extract the following lemma.

Lemma 1 *If for any* $\delta > 0$

$$\liminf_{N\to\infty} \inf_{\|\theta-\bar{\theta}\|\geq\delta} [S_N(\theta) - S_N(\bar{\theta})] > 0 \text{ a.s.} \tag{11}$$

then $\hat{\theta}^N_{LS} \overset{\text{a.s.}}{\to} \bar{\theta}$ *as* $N \to \infty$. *If for any* $\delta > 0$

$$\Pr\left\{ \inf_{\|\theta-\bar{\theta}\|\geq\delta} [S_N(\theta) - S_N(\bar{\theta})] > 0 \right\} \to 1\,,\ N \to \infty\,, \tag{12}$$

then $\hat{\theta}^N_{LS} \overset{\text{p}}{\to} \bar{\theta}$ *as* $N \to \infty$.

We can then prove the convergence of the LS estimator (in probability and a.s.) when the sum $\sum_{k=1}^N [\eta(x_k,\theta) - \eta(x_k,\bar{\theta})]^2$ tends to infinity fast enough for $\|\theta - \bar{\theta}\| \geq \delta > 0$ and the design space $\mathcal{X}$ for the x_k's is finite.

Theorem 1 *Let* $\{x_i\}$ *be a design sequence* on a finite set $\mathcal{X}$. *If* $D_N(\theta,\bar{\theta}) = \sum_{k=1}^N [\eta(x_k,\theta) - \eta(x_k,\bar{\theta})]^2$ *satisfies*

$$\text{for all } \ \delta > 0\,,\ \left[\inf_{\|\theta-\bar{\theta}\|\geq\delta} D_N(\theta,\bar{\theta}) \right] /(\log\log N) \to \infty\,,\ N \to \infty\,, \tag{13}$$

then $\hat{\theta}^N_{LS} \overset{\text{a.s.}}{\to} \bar{\theta}$ *as* $N \to \infty$. *If* $D_N(\theta,\bar{\theta})$ *simply satisfies*

$$\text{for all } \ \delta > 0\,,\ \inf_{\|\theta-\bar{\theta}\|\geq\delta} D_N(\theta,\bar{\theta}) \to \infty \ \text{as} \ \ N \to \infty\,, \tag{14}$$

then $\hat{\theta}^N_{LS} \overset{\text{p}}{\to} \bar{\theta}$, $N \to \infty$.

Proof. The proof is based on Lemma 1. We have

$$\begin{aligned} S_N(\theta) - S_N(\bar{\theta}) &= D_N(\theta,\bar{\theta}) \left[1 + 2\, \frac{\sum_{x\in\mathcal{X}} \left(\sum_{k=1,\, x_k=x}^N \varepsilon_k \right) [\eta(x,\bar{\theta}) - \eta(x,\theta)]}{D_N(\theta,\bar{\theta})} \right] \\ &\geq D_N(\theta,\bar{\theta}) \left[1 - 2\, \frac{\sum_{x\in\mathcal{X}} \left| \sum_{k=1,\, x_k=x}^N \varepsilon_k \right| |\eta(x,\bar{\theta}) - \eta(x,\theta)|}{D_N(\theta,\bar{\theta})} \right] . \end{aligned}$$

From Lemma 1, under the condition (13) it suffices to prove that

$$\sup_{\|\theta-\bar{\theta}\|\geq\delta} \frac{\sum_{x\in\mathcal{X}} \left|\sum_{k=1,\, x_k=x}^{N} \varepsilon_k\right| |\eta(x,\bar{\theta})-\eta(x,\theta)|}{D_N(\theta,\bar{\theta})} \overset{\text{a.s.}}{\to} 0 \qquad (15)$$

for any $\delta > 0$ to obtain the strong consistency of $\hat{\theta}^N_{LS}$. Since $D_N(\theta,\bar{\theta}) \to \infty$ and $\mathcal{X}$ is finite, only the design points such that $r_N(x) \to \infty$ have to be considered, where $r_N(x)$ denotes the number of times x appears in the sequence $x_1,\ldots,x_N$. Define $\beta(n) = \sqrt{n \log\log n}$. From the law of the iterated logarithm,

$$\text{for all } \ x\in\mathcal{X}\,, \ \limsup_{r_N(x)\to\infty} \left| \frac{1}{\beta[r_N(x)]} \sum_{k=1,\, x_k=x}^{N} \varepsilon_k \right| = \sigma\sqrt{2}\,, \quad \text{almost surely}\,. \qquad (16)$$

Moreover, $D_N(\theta,\bar{\theta}) \geq D_N^{1/2}(\theta,\bar{\theta})\sqrt{r_N(x)}|\eta(x,\bar{\theta})-\eta(x,\theta)|$ for any $x\in\mathcal{X}$, so that

$$\frac{\beta[r_N(x)]|\eta(x,\bar{\theta})-\eta(x,\theta)|}{D_N(\theta,\bar{\theta})} \leq \frac{[\log\log r_N(x)]^{1/2}}{D_N^{1/2}(\theta,\bar{\theta})}\,.$$

Therefore,

$$\frac{\left|\sum_{k=1,\, x_k=x}^{N} \varepsilon_k\right| |\eta(x,\bar{\theta})-\eta(x,\theta)|}{D_N(\theta,\bar{\theta})} \leq \left| \frac{\sum_{k=1,\, x_k=x}^{N} \varepsilon_k}{\beta[r_N(x)]} \right| \frac{[\log\log r_N(x)]^{1/2}}{D_N^{1/2}(\theta,\bar{\theta})}\,,$$

which, together with (13) and (16), gives (15).

When $\inf_{\|\theta-\bar{\theta}\|\geq\delta} D_N(\theta,\bar{\theta}) \to \infty$ as $N\to\infty$, we only need to prove that

$$\sup_{\|\theta-\bar{\theta}\|\geq\delta} \frac{\sum_{x\in\mathcal{X}} \left|\sum_{k=1,\, x_k=x}^{N} \varepsilon_k\right| |\eta(x,\bar{\theta})-\eta(x,\theta)|}{D_N(\theta,\bar{\theta})} \overset{\text{p}}{\to} 0 \qquad (17)$$

for any $\delta > 0$ to obtain the weak consistency of $\hat{\theta}^N_{LS}$. We proceed as above and only consider the design points such that $r_N(x) \to \infty$, with now $\beta(n) = \sqrt{n}$. From the central limit theorem, for any $x \in \mathcal{X}$, $\left(\sum_{k=1,\, x_k=x}^{N} \varepsilon_k\right)/\sqrt{r_N(x)} \overset{\text{d}}{\to} \omega_x \sim \mathcal{N}(0,\sigma^2)$ as $r_N(x)\to\infty$ and is thus bounded in probability. Also, for any $x\in\mathcal{X}$, $\sqrt{r_N(x)}|\eta(x,\bar{\theta})-\eta(x,\theta)|/D_N(\theta,\bar{\theta}) \leq D_N^{-1/2}(\theta,\bar{\theta})$, so that (14) implies (17). ■

When the design space $\mathcal{X}$ is finite one can thus invoke Theorem 1 to ensure the consistency of $\hat{\theta}^N_{LS}$. Regular asymptotic normality then follows for suitable functions $h(\cdot)$.

Theorem 2 *Let $\{x_i\}$ be a design sequence on a finite set $\mathcal{X}$, with the property that the associated empirical measure (strongly) converges to ξ (possibly singular), that is,* $\lim_{N\to\infty} r_N(x)/N = \xi(x)$ *for any $x\in\mathcal{X}$, with $r_N(x)$ the number of* *times x appears in the sequence $x_1, \ldots, x_N$. Suppose that the assumptions $\mathbf{H1}_\eta$, $\mathbf{H2}_\eta$ and $\mathbf{H}_h$ are satisfied, with $\partial h(\theta)/\partial\theta|_{\bar\theta} \neq \mathbf{0}$, and that $D_N(\theta, \bar\theta)$ satisfies (13). Then,*

$$\frac{\partial h(\theta)}{\partial \theta}\Big|_{\bar\theta} \in \mathcal{M}[\mathbf{M}(\xi, \bar\theta)]\,, \tag{18}$$

implies that $h(\hat\theta^N_{LS})$ satisfies the regular asymptotic normality property (9), where the choice of the g-inverse is arbitrary.

Proof. Since $\hat\theta^N_{LS} \overset{\text{a.s.}}{\to} \bar\theta \in \text{int}(\Theta)$, there exists N_0 such that $\hat\theta^N_{LS}$ is in some convex neighborhood of $\bar\theta$ for all N larger than N_0 and, for all $i = 1, \ldots, p = \dim(\theta)$, a Taylor development of the i-th component of the gradient of the LS criterion (3) gives

$$\{\nabla_\theta S_N(\hat\theta^N_{LS})\}_i = 0 = \{\nabla_\theta S_N(\bar\theta)\}_i + \{\nabla^2_\theta S_N(\beta^N_i)(\hat\theta^N_{LS} - \bar\theta)\}_i\,, \tag{19}$$

with β^N_i between $\hat\theta^N_{LS}$ and $\bar\theta$ (and β^N_i measurable, see [3]). Using the fact that $\mathcal{X}$ is finite we obtain $\nabla_\theta S_N(\bar\theta)/\sqrt{N} \overset{\text{d}}{\to} \mathbf{v} \sim \mathcal{N}(\mathbf{0}, 4\mathbf{M}(\xi, \bar\theta))$ and $\nabla^2_\theta S_N(\beta^N_i)/N \overset{\text{a.s.}}{\to} 2\mathbf{M}(\xi, \bar\theta)$ as $N \to \infty$. Combining this with (19), we get

$$\sqrt{N}\mathbf{c}^\top \mathbf{M}(\xi, \bar\theta)(\hat\theta^N_{LS} - \bar\theta) \overset{\text{d}}{\to} z \sim \mathcal{N}(0, \mathbf{c}^\top \mathbf{M}(\xi, \bar\theta)\mathbf{c})\,,\ N \to \infty\,,$$

for any $\mathbf{c} \in \mathbb{R}^p$. Applying the Taylor formula again we can write

$$\sqrt{N}[h(\hat\theta^N_{LS}) - h(\bar\theta)] = \sqrt{N}\,\frac{\partial h(\theta)}{\partial \theta^\top}\Big|_{\alpha^N}(\hat\theta^N_{LS} - \bar\theta)$$

for some α^N between $\hat\theta^N_{LS}$ and $\bar\theta$ and $\partial h(\theta)/\partial\theta\big|_{\alpha^N} \overset{\text{a.s.}}{\to} \partial h(\theta)/\partial\theta\big|_{\bar\theta}$ as $N \to \infty$. When (18) is satisfied we can write $\partial h(\theta)/\partial\theta\big|_{\bar\theta} = \mathbf{M}(\xi, \bar\theta)\mathbf{u}$ for some $\mathbf{u} \in \mathbb{R}^p$, which gives (9). ■

Notice that when $\mathbf{M}(\xi, \bar\theta)$ has full rank the condition (18) is automatically satisfied so that the other conditions of Theorem 2 are sufficient for the asymptotic normality (8). The conclusion of the Theorem remains valid when $D_N(\theta, \bar\theta)$ only satisfies (14) (convergence in probability of $\hat\theta^N_{LS}$) with Θ a convex set, see, e.g., [1, Th. 4.2.2].

3 Properties of standard regularization

Consider a regularized version of the c-optimality criterion defined by

$$\Phi^\gamma_c(\mathbf{M}) = \Phi_c[(1-\gamma)\mathbf{M} + \gamma\tilde{\mathbf{M}}]$$

with $\Phi_c(\cdot)$ given by (10), γ a small positive number and $\tilde{\mathbf{M}}$ a fixed nonsingular $p \times p$ matrix of $\mathbb{M}^{\geq}$. From the linearity of $\mathbf{M}(\xi, \theta^0)$ in ξ, when $\tilde{\mathbf{M}} = \mathbf{M}(\tilde\xi, \theta^0)$ with $\tilde\xi$ nonsingular this equivalently defines the criterion

$$\phi^\gamma_c(\xi) = \phi_c[(1-\gamma)\xi + \gamma\tilde\xi]$$

with $\phi_c(\xi) = \Phi_c[\mathbf{M}(\xi, \theta^0)]$. Let ξ^* and ξ_γ^* be two measures respectively optimal for $\phi_c(\cdot)$ and $\phi_c^\gamma(\cdot)$. We have $\phi_c(\xi^*) \leq \phi_c[(1-\gamma)\xi_\gamma^* + \gamma\tilde{\xi}] = \phi_c^\gamma(\xi_\gamma^*) \leq \phi_c[(1-\gamma)\xi^* + \gamma\tilde{\xi}] \leq (1-\gamma)\phi_c(\xi^*) + \gamma\phi_c(\tilde{\xi})$, where the last inequality follows from the convexity of $\phi_c(\cdot)$. Therefore,

$$0 \leq \phi_c^\gamma(\xi_\gamma^*) - \phi_c(\xi^*) \leq \gamma[\phi_c(\tilde{\xi}) - \phi_c(\xi^*)]$$

which tends to zero as $\gamma \to 0$, showing that $\hat{\xi}_\gamma = (1-\gamma)\xi_\gamma^* + \gamma\tilde{\xi}$ tends to be c-optimal when γ decreases to zero.

We emphasize that c-optimality is defined for $\theta^0 \neq \bar{\theta}$. Let $x^{(1)}, \ldots, x^{(s)}$ be the support points of a c-optimal measure ξ^*, complement them by $x^{(s+1)}, \ldots, x^{(s+k)}$ so that the measure $\tilde{\xi}$ supported at $x^{(1)}, \ldots, x^{(s+k)}$ (with, e.g., equal weight at each point) is nonsingular. When N observations are made, to the measure $(1-\gamma)\xi^* + \gamma\tilde{\xi}$ corresponds a design that places approximately $\gamma N/(s+k)$ observations at each of the points $x^{(s+1)}, \ldots, x^{(s+k)}$. The example below shows that the speed of convergence of $\mathbf{c}^\top\hat{\theta}_{LS}^N$ to $\mathbf{c}^\top\bar{\theta}$ may be arbitrarily slow when γ tends to zero, thereby contradicting the acceptance of ξ^* as a c-optimal design for $\bar{\theta}$.

Example: Consider the regression model defined by (1,2) with

$$\eta(x, \theta) = \frac{\theta_1}{\theta_1 - \theta_2}\left[\exp(-\theta_2 x) - \exp(-\theta_1 x)\right],$$

$\mathcal{X} = [0, 10]$ and $\sigma^2 = 1$. The D-optimal design measure ξ_D^* on $\mathcal{X}$ maximizing $\log\det\mathbf{M}(\xi, \theta^0)$ for the nominal parameters $\theta^0 = (0.7,\ 0.2)^\top$ puts mass 1/2 at each of the two support points given approximately by $x^{(1)} = 1.25$, $x^{(2)} = 6.60$.

Figure 1 shows the set $\{\mathbf{f}_{\theta^0}(x) : x \in \mathcal{X}\}$ (solid line), its symmetric $\{-\mathbf{f}_{\theta^0}(x) : x \in \mathcal{X}\}$ (dashed line) and their convex closure $\mathcal{F}_{\theta^0}$, called the Elfving set (shaded region), together with the minimum-volume ellipsoid containing $\mathcal{F}_{\theta^0}$ (the points of contact with $\mathcal{F}_{\theta^0}$ correspond to the support points of ξ_D^*).

From Elfving's theorem [2], when $x^* \in [x^{(1)}, x^{(2)}]$ the c-optimal design minimizing $\mathbf{c}^\top\mathbf{M}^{-1}(\xi, \theta^0)\mathbf{c}$ with $\mathbf{c} = \beta\mathbf{f}_{\theta^0}(x_*)$, $\beta \neq 0$, is the delta measure δ_{x_*}. Obviously, the singular design δ_{x_*} only allows us to estimate $\eta(x_*, \theta)$ and not $h(\theta) = \mathbf{c}^\top\theta$.

Select now a second design point $x^0 \neq x_*$ and suppose that when N observations are performed at the design points $x_1, \ldots, x_N$, m of them coincide with x^0 and $N-m$ with x_*, where $m/(\log\log N) \to \infty$ with $m/N \to 0$. Then, for $x^0 \neq 0$ the conditions of Theorem 1 are satisfied. Indeed, the design space equals $\{x^0, x_*\}$ and is thus finite, and

$$\begin{aligned}
D_N(\theta, \bar{\theta}) &= \sum_{k=1}^{N}[\eta(x_k, \theta) - \eta(x_k, \bar{\theta})]^2 \\
&= (N-2m)[\eta(x_*, \theta) - \eta(x_*, \bar{\theta})]^2 \\
&\quad + m\left\{[\eta(x_*, \theta) - \eta(x_*, \bar{\theta})]^2 + [\eta(x^0, \theta) - \eta(x^0, \bar{\theta})]^2\right\}
\end{aligned}$$

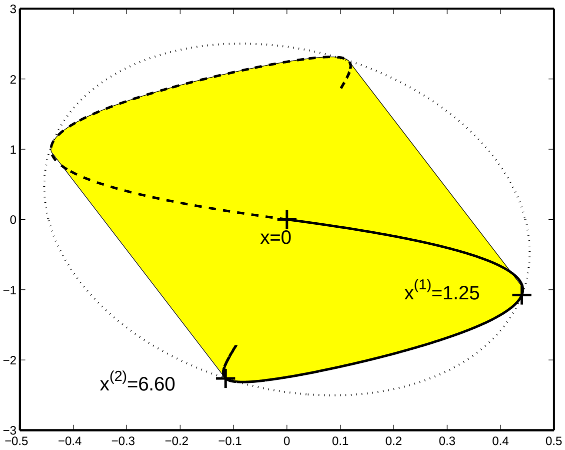

Figure 1: Elfving set.

so that $\inf_{\|\theta-\bar{\theta}\|>\delta} D_N(\theta,\bar{\theta}) \geq mC(x^0,x_*,\delta)$, with $C(x^0,x_*,\delta)$ a positive constant, and $\inf_{\|\theta-\bar{\theta}\|>\delta} D_N(\theta,\bar{\theta})/(\log\log N) \to \infty$ as $N \to \infty$. Therefore, although the empirical measure ξ_N of the design points in the experiment converges strongly to the singular design δ_{x_*}, this convergence is sufficiently slow to make $\hat{\theta}^N_{LS}$ (strongly) consistent. Moreover, for $h(\cdot)$ a function satisfying the conditions of Theorem 2, $h(\hat{\theta}^N_{LS})$ satisfies the regular asymptotic property (9). In the present situation, this means that when $\partial h(\theta)/\partial\theta\big|_{\bar{\theta}} = \beta \mathbf{f}_{\bar{\theta}}(x_*)$ for some $\beta \in \mathbb{R}$, then $\sqrt{N}[h(\hat{\theta}^N_{LS}) - h(\bar{\theta})] \xrightarrow{\text{d}} \omega \sim \mathcal{N}(0, \left[\partial h(\theta)/\partial\theta^\top\, \mathbf{M}^-(\delta_{x_*},\theta)\, \partial h(\theta)/\partial\theta\right]_{\bar{\theta}})$. This holds for instance when $h(\cdot) = \eta(x_*,\cdot)$ (or is a function of $\eta(x_*,\cdot)$).

There is, however, a severe limitation in the application of this result in practical situations. Indeed, the direction $\mathbf{f}_{\bar{\theta}}(x_*)$ for which regular asymptotic normality holds is unknown since $\bar{\theta}$ is unknown. Let $\mathbf{c}$ be a given direction of interest, the associated c-optimal design ξ^* is determined for the nominal value θ^0. For instance, when $\mathbf{c} = (0,\ 1)^\top$ (which means that one is only interested in the estimation of the component θ_2), $\xi^* = \delta_{x_*}$ with x_* solution of $\{\mathbf{f}_{\theta^0}(x)\}_1 = 0$ (see Figure 1), that is, x_* satisfies

$$\theta^0_2 = [\theta^0_2 + \theta^0_1(\theta^0_1 - \theta^0_2)x_*]\, \exp[-(\theta^0_1 - \theta^0_2)x_*]\,. \tag{20}$$

For $\theta^0 = (0.7,\ 0.2)^\top$, this gives $x_* = x_*(\theta^0) \simeq 4.28$. In general, $\mathbf{f}_{\bar{\theta}}(x_*) \neq \mathbf{f}_{\theta^0}(x_*)$ to which $\mathbf{c}$ is proportional. Therefore, $\mathbf{c} \notin \mathcal{M}[\mathbf{M}(\xi^*,\bar{\theta})]$ and regular asymptotic normality does not hold for $\mathbf{c}^\top\hat{\theta}^N_{LS}$.

The example is simple enough to be able to investigate the limiting behavior of $\mathbf{c}^\top\hat{\theta}^N_{LS}$ by direct calculation. A Taylor development of the LS criterion $S_N(\theta)$ gives (19) where $\beta^N_i \overset{\text{a.s.}}{\to} \bar{\theta}$ as $N \to \infty$, $i = 1,2$. Direct calculations give

$$\begin{aligned}
\nabla_\theta S_N(\bar{\theta}) &= -2\left[\sqrt{m}\,\beta_m\, \mathbf{f}_{\bar{\theta}}(x^0) + \sqrt{N-m}\,\gamma_{N-m}\, \mathbf{f}_{\bar{\theta}}(x_*)\right]\,, \\
\nabla^2_\theta S_N(\bar{\theta}) &= 2\left[m\, \mathbf{f}_{\bar{\theta}}(x^0)\mathbf{f}^\top_{\bar{\theta}}(x^0) + (N-m)\, \mathbf{f}_{\bar{\theta}}(x_*)\mathbf{f}^\top_{\bar{\theta}}(x_*)\right] + \mathcal{O}_p(\sqrt{m})\,,
\end{aligned}$$

where $\beta_m = (1/\sqrt{m}) \sum_{x_i = x^0} \varepsilon_i$ and $\gamma_{N-m} = (1/\sqrt{N-m}) \sum_{x_i = x_*} \varepsilon_i$ are independent random variables that tend to be distributed $\mathcal{N}(0,1)$ as $m \to \infty$ and $N - m \to \infty$. We then obtain,

$$\hat{\theta}^N_{LS} - \bar{\theta} = \frac{1}{\Delta(x_*, x^0)} \left\{ \frac{\gamma_{N-m}}{\sqrt{N-m}} \begin{pmatrix} \{\mathbf{f}_{\bar{\theta}}(x^0)\}_2 \\ -\{\mathbf{f}_{\bar{\theta}}(x^0)\}_1 \end{pmatrix} + \frac{\beta_m}{\sqrt{m}} \begin{pmatrix} -\{\mathbf{f}_{\bar{\theta}}(x_*)\}_2 \\ \{\mathbf{f}_{\bar{\theta}}(x_*)\}_1 \end{pmatrix} \right\} + o_p(1/\sqrt{m}) \,,$$

where $\Delta(x_*, x^0) = \det(\mathbf{f}_{\bar{\theta}}(x_*), \mathbf{f}_{\bar{\theta}}(x^0))$. Therefore, $\sqrt{N}\mathbf{f}^\top_{\bar{\theta}}(x_*)(\hat{\theta}^N_{LS} - \bar{\theta})$ is asymptotically normal $\mathcal{N}(0,1)$ whereas for any direction $\mathbf{c}$ not parallel to $\mathbf{f}_{\bar{\theta}}(x_*)$ and not orthogonal to $\mathbf{f}_{\bar{\theta}}(x^0)$, $\sqrt{m}\mathbf{c}^\top(\hat{\theta}^N_{LS} - \bar{\theta})$ is asymptotically normal (and $\mathbf{c}^\top(\hat{\theta}^N_{LS} - \bar{\theta})$ converges not faster than $1/\sqrt{m}$). In particular, $\sqrt{m}\mathbf{f}^\top_{\bar{\theta}}(x^0)(\hat{\theta}^N_{LS} - \bar{\theta})$ is asymptotically normal $\mathcal{N}(0,1)$ and $\sqrt{m}\{\hat{\theta}^N_{LS} - \bar{\theta}\}_2$ is asymptotically normal $\mathcal{N}(0, \{\mathbf{f}_{\bar{\theta}}(x_*)\}^2_1 / \Delta^2(x_*, x^0))$. $\square$

The previous example has illustrated that letting γ tend to zero in a regularized c-optimal design $(1-\gamma)\xi^* + \gamma\tilde{\xi}$ raises important difficulties (one may refer to [8] for an example with a linear model and a nonlinear function $h(\theta)$). We shall therefore consider γ as fixed in what follows. It is interesting, nevertheless, to investigate the behavior of the c-optimality criterion when the regularized measure $(1-\gamma)\xi^* + \gamma\tilde{\xi}$ approaches ξ^* in some sense. Since γ is now fixed, we let the support points of $\tilde{\xi}$ approach those of ξ^*. This is illustrated by continuing the example above.

Example (continued): Place the proportion $m = N/2$ of the observations at x^0 and consider the design measure $\xi_{\gamma,x^0} = (1-\gamma)\delta_{x_*} + \gamma\delta_{x^0}$ with $\gamma = 1/2$. Since the c-optimal design is δ_{x_*}, we consider the limiting behavior of $\mathbf{c}^\top(\hat{\theta}^N_{LS} - \bar{\theta})$ when N tends to infinity for x^0 approaching x_*. The nonsingularity of $\xi_{1/2,x^0}$ for $x^0 \neq x_*$ (and $x^0 \neq 0$) implies that $\sqrt{N}\mathbf{c}^\top(\hat{\theta}^N_{LS} - \bar{\theta})$ is asymptotically normal $\mathcal{N}(0, \mathbf{c}^\top\mathbf{M}^{-1}(\xi_{1/2,x^0}, \bar{\theta})\mathbf{c})$.

The asymptotic variance $\mathbf{c}^\top\mathbf{M}^{-1}(\xi_{1/2,x^0}, \bar{\theta})\mathbf{c}$ tends to infinity as x^0 tends to x_* when $\mathbf{c}$ is not proportional to $\mathbf{f}_{\bar{\theta}}(x_*)$, see Figure 2. Take $\mathbf{c} = \mathbf{f}_{\bar{\theta}}(x_*)$. Then, $\mathbf{f}^\top_{\bar{\theta}}(x_*)\mathbf{M}^{-1}(\xi_{1/2,x^0}, \bar{\theta})\mathbf{f}_{\bar{\theta}}(x_*)$ equals 2 for any $x^0 \neq x_*$, twice more than what could be achieved with the singular design δ_{x_*} since $\mathbf{f}^\top_{\bar{\theta}}(x_*)\mathbf{M}^{-}(\delta_{x_*}, \bar{\theta})\mathbf{f}_{\bar{\theta}}(x_*) = 1$ (this result is similar to that in [6, p. 67] and is caused by the fact that $\Phi_c(\cdot)$ is only semi-continuous at a singular $\mathbf{M}$). $\square$

The example above shows that not all regularizations are legitimate: the regularized design should be close to the optimal one ξ^* in some suitable sense in order to avoid the discontinuity of $\Phi_c(\cdot)$ at a singular $\mathbf{M}$.

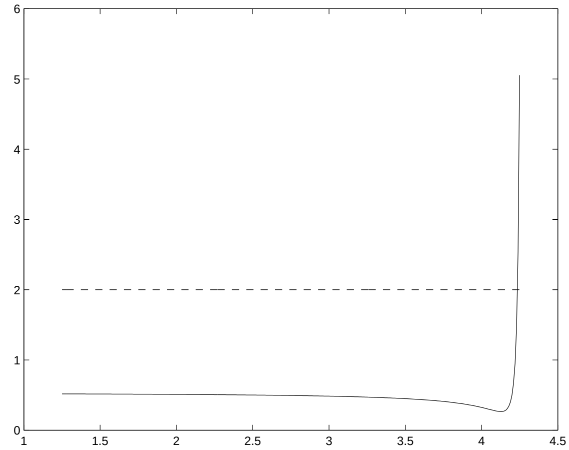

Figure 2: $\mathbf{c}^\top \mathbf{M}^{-1}(\xi_{1/2,x^0}, \bar{\theta})\mathbf{c}$ (solid line) and $\mathbf{f}_{\bar{\theta}}^\top(x_*)\mathbf{M}^{-1}(\xi_{1/2,x^0}, \bar{\theta})\mathbf{f}_{\bar{\theta}}(x_*)$ (dashed line) for x^0 varying between 1.25 and $x_* = 4.28$; $\bar{\theta} = (0.65,\ 0.25)^\top$, $\theta^0 = (0.7,\ 0.2)^\top$ and $\mathbf{c} = (0,\ 1)^\top$ (so that δ_{x_*} is c-optimal for $\mathbf{c}$ and θ^0).

4 Minimax regularization

4.1 Estimation of a nonlinear function of θ

Consider first the case where the function of interest $h(\theta)$ is nonlinear in θ. We should then ideally take $\mathbf{c}_{\bar{\theta}} = \partial h(\theta)/\partial\theta\big|_{\bar{\theta}}$ in the definition of the optimality criterion. However, since $\bar{\theta}$ is unknown, a direct application of local c-optimal design consists in using the direction $\mathbf{c}_{\theta^0} = \partial h(\theta)/\partial\theta\big|_{\theta^0}$, with the risk that θ and $h(\theta)$ are not estimable from the associated optimal design ξ^* if it is singular. One can then consider instead a set Θ^0 (a finite set or a compact subset of $\mathbb{R}^p$) of possible values for $\bar{\theta}$ around θ^0 in the definition of the directions of interest, and the associated c-minimax optimality criterion becomes

$$\phi_{\mathcal{C}}(\xi) = \max_{\theta\in\Theta^0} \mathbf{c}_\theta^\top \mathbf{M}^-(\xi, \theta^0)\mathbf{c}_\theta\,, \tag{21}$$

or equivalently $\phi_{\mathcal{C}}(\xi) = \max_{\mathbf{c}\in\mathcal{C}} \mathbf{c}^\top \mathbf{M}^-(\xi, \theta^0)\mathbf{c}$ with $\mathcal{C} = \{\mathbf{c}_\theta : \theta \in \Theta^0\}$. A measure $\xi^*(\mathcal{C})$ on $\mathcal{X}$ that minimizes $\phi_{\mathcal{C}}(\xi)$ is said to be (locally) c-minimax optimal. When $\mathcal{C}$ is large enough (in particular when the vectors in $\mathcal{C}$ span $\mathbb{R}^p$), $\xi^*(\mathcal{C})$ is nonsingular. According to Theorem 2, a design sequence on a finite set $\mathcal{X}$ (containing the support of $\xi^*(\mathcal{C})$) such that the associated empirical measure converges strongly to $\xi^*(\mathcal{C})$ then ensures the asymptotic normality property (8).

4.2 Estimation of a linear function of θ: regularization via D-optimal design

When the function of interest is $h(\theta) = \mathbf{c}^\top\theta$ with the direction $\mathbf{c}$ fixed, the construction of an admissible set $\mathcal{C}$ of directions for c-minimax optimal design is somewhat artificial and a specific procedure is required. The rest of the section is devoted to this situation. The first approach presented is based on D-optimality and applies when the c-optimal measure is a one-point measure.

Define a (local) c-maximin efficient measure ξ^*_{mm} for $\mathcal{C}$ as a measure on $\mathcal{X}$ that maximizes

$$E_{mm}(\xi) = \min_{\mathbf{c}\in\mathcal{C}} \frac{\mathbf{c}^\top \mathbf{M}^-[\xi^*(\mathbf{c}), \theta^0]\mathbf{c}}{\mathbf{c}^\top \mathbf{M}^-(\xi, \theta^0)\mathbf{c}},$$

with $\xi^*(\mathbf{c})$ a c-optimal design measure minimizing $\mathbf{c}^\top \mathbf{M}^-(\xi, \theta^0)\mathbf{c}$. When the c-optimal design $\xi^*(\mathbf{c})$ is the delta measure δ_{x_*} it seems reasonable to consider measures that are supported in the neighborhood of x_*. One may then use the following result of Kiefer [4] to obtain a c-maximin efficient measure through D-optimal design.

Theorem 3 *A design measure ξ^*_{mm} on $\mathcal{X}$ is c-maximin efficient for $\mathcal{C}_\mathcal{X} = \{\mathbf{f}_{\theta^0}(x) : x \in \mathcal{X}\}$ if and only if it is D-optimal on $\mathcal{X}$, that is, it maximizes $\log\det \mathbf{M}(\xi, \theta^0)$.*

The construction is as follows. Define

$$\mathcal{X}_\delta = \mathcal{B}(x_*, \delta) \cap \mathcal{X}, \quad (22)$$

with $\mathcal{B}(x_*, \delta)$ the ball of centre x_* and radius δ in $\mathbb{R}^d$, and define $\mathcal{C}_\delta = \{\mathbf{f}_{\theta^0}(x) : x \in \mathcal{X}_\delta\}$. From Theorem 3, a measure ξ^*_δ is c-maximin efficient for $\mathbf{c} \in \mathcal{C}_\delta$ if and only if it is D-optimal on $\mathcal{X}_\delta$. Suppose that $\mathcal{C}_\delta$ spans $\mathbb{R}^p$ when $\delta > 0$, the measure ξ^*_δ is then non singular for $\delta > 0$ (with $\xi^*_0 = \xi^*(\mathbf{c})$). Various values of δ are associated with different designs ξ^*_δ. One may then choose δ by minimizing

$$J(\delta) = \max_{\theta\in\Theta^0} \Phi_c[\mathbf{M}(\xi^*_\delta, \theta)], \quad (23)$$

where Θ^0 defines a feasible set for the unknown parameter vector $\bar{\theta}$. Each evaluation of $J(\delta)$ requires the determination of a D-optimal design on a set $\mathcal{X}_\delta$ and the determination of the minimum with respect to $\theta \in \Theta^0$, but the D-optimal design is often easily obtained, see the example below, and the set Θ^0 can be discretized to facilitate the determination of the maximum.

Example (continued): Take $\mathbf{c} = (0\,,\ 1)^\top$ and $\theta^0 = (0.7\,,\ 0.2)^\top$. Choosing $\mathcal{X}_\delta$ as in (22) gives $\mathcal{C}_\delta = \{\mathbf{f}_{\theta^0}(x) : x \in [x_* - \delta, x_* + \delta]\}$, with $x_* \simeq 4.28$, and the corresponding c-maximin efficient measure is $\xi^*_\delta = (1/2)\delta_{x_*-\delta} + (1/2)\delta_{x_*+\delta}$. Fig. 3 shows $\mathbf{c}^\top \mathbf{M}^{-1}(\xi^*_\delta, \bar{\theta})\mathbf{c}$ and $\mathbf{f}^\top_{\bar{\theta}}(x_*)\mathbf{M}^{-1}(\xi^*_\delta, \bar{\theta})\mathbf{f}_{\bar{\theta}}(x_*)$ as functions of δ. Notice that $\mathbf{f}^\top_{\bar{\theta}}(x_*)\mathbf{M}^{-1}(\xi^*_\delta, \bar{\theta})\mathbf{f}_{\bar{\theta}}(x_*)$ tends to 1 as δ tends to zero, indicating that the form of the neighborhood used in the construction of $\mathcal{X}_\delta$ has a strong influence on the performance of ξ^*_δ (in terms of c-optimality) when δ tends to zero. Indeed, taking $\mathcal{X}_\delta = [x^0, x_*]$ with $x^0 = x_* - \delta$ yields the same situation as that depicted in Fig. 2.

The curve showing $\mathbf{c}^\top \mathbf{M}^{-1}(\xi^*_\delta, \bar{\theta})\mathbf{c}$ in Fig. 3 indicates the presence of a minimum around $\delta = 0.5$. Fig. 4 presents $J(\delta)$ given by (23) as a function of δ when $\Theta^0 = [0.6, 0.8] \times [0.1, 0.3]$, indicating a minimum around $\delta = 1.45$ (the maximum over θ is attained at the endpoints $\theta_1 = 0.8, \theta_2 = 0.3$ for any δ). □

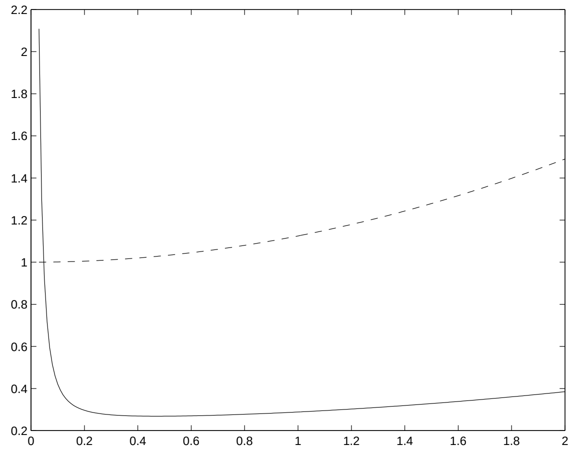

Figure 3: $\mathbf{c}^\top \mathbf{M}^{-1}(\xi_\delta^*, \bar{\theta})\mathbf{c}$ (solid line) and $\mathbf{f}_{\bar{\theta}}^\top(x_*)\mathbf{M}^{-1}(\xi_\delta^*, \bar{\theta})\mathbf{f}_{\bar{\theta}}(x_*)$ (dashed line) for δ between 0 and 2; $x_* = 4.28$, $\bar{\theta} = (0.65,\ 0.25)^\top$, $\theta^0 = (0.7,\ 0.2)^\top$ and $\mathbf{c} = (0,\ 1)^\top$ (so that δ_{x_*} is c-optimal for $\mathbf{c}$ and θ^0).

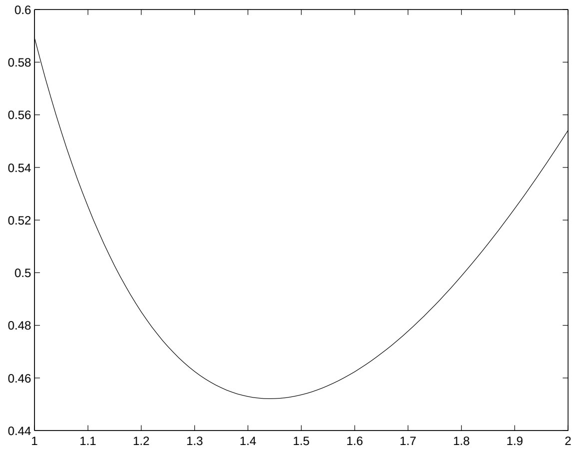

Figure 4: $\max_{\theta \in \Theta^0} \mathbf{c}^\top \mathbf{M}^{-1}(\xi_\delta^*, \theta)\mathbf{c}$ as a function of $\delta \in [1, 2]$ for $\Theta^0 = [0.6, 0.8] \times [0.1, 0.3]$.

5 Regularization by combination of c-optimal designs

We say that $h(\theta)$ is *locally estimable* at θ for the design ξ in the regression model (1,2) if the condition (7) is locally satisfied, that is, if there exists a neighborhood Θ_θ of θ such that

$$\forall \theta' \in \Theta_\theta \,, \; \int_{\mathcal{X}} [\eta(x,\theta') - \eta(x,\theta)]^2 \, \xi(dx) = 0 \Rightarrow h(\theta') = h(\theta) \,. \qquad (24)$$

Consider again the case of a linear function of interest $h(\theta) = \mathbf{c}^\top\theta$ with the direction $\mathbf{c}$ fixed. The next theorem indicates that when $\mathbf{c}^\top\theta$ is not (locally) estimable at θ^0 from the c-optimal design ξ^* it means that the support of ξ^* depends on the value θ^0 for which it is calculated. By combining different c-optimal designs obtained at various nominal values $\theta^{0,i}$ one can thus easily construct a nonsingular design from which θ, and thus $\mathbf{c}^\top\theta$, can be estimated. When the true value of $\bar{\theta}$ is not too far from the $\theta^{0,i}$'s, this design will be almost c-optimal for $\bar{\theta}$.

Theorem 4 *Consider a linear function of interest $h(\theta) = \mathbf{c}^\top\theta$, $\mathbf{c} \neq \mathbf{0}$, in a regression model (1,2) satisfying the assumptions $\mathbf{H1}_\eta$, $\mathbf{H2}_\eta$ and $\mathbf{H}_h$. Let $\xi^* = \xi^*(\theta^0)$ be a (local) c-optimal design minimizing $\mathbf{c}^\top\mathbf{M}^-(\xi,\theta^0)\mathbf{c}$. Then, $h(\theta)$ being not locally estimable for ξ^* at θ^0 implies that the support of ξ^* varies with the choice of θ^0.*

Proof. The proof is by contradiction. Suppose that the support of $\xi^*(\theta)$ does not depend on θ. We show that it implies that $h(\theta)$ is locally estimable at θ for ξ^*.

Suppose, without any loss of generality, that $\mathbf{c} = (c_1,\ldots,c_p)^\top$ with $c_1 \neq 0$ and consider the reparametrization defined by $\beta = (\mathbf{c}^\top\theta, \theta_2, \ldots, \theta_p)^\top$, so that $\theta = \theta(\beta) = \mathbf{J}\beta$ with $\mathbf{J}$ the (jacobian) matrix

$$\mathbf{J} = \begin{pmatrix} 1/c_1 & -\mathbf{c}'^\top/c_1 \\ \mathbf{0}_{p-1} & \mathbf{I}_{p-1} \end{pmatrix} ,$$

where $\mathbf{c}' = (c_2,\ldots,c_p)^\top$ and $\mathbf{0}_{p-1}$, $\mathbf{I}_{p-1}$ respectively denote the $(p-1)$-dimensional null vector and identity matrix. From Elfving's Theorem,

$$\int_{\mathcal{S}^*} \frac{\partial\eta(x,\theta)}{\partial\theta} \xi^*(dx) - \int_{\mathcal{S}_{\xi^*}\setminus\mathcal{S}^*} \frac{\partial\eta(x,\theta)}{\partial\theta} \xi^*(dx) = \gamma\mathbf{c}$$

with $\gamma = \gamma(\theta) > 0$, $\mathcal{S}_{\xi^*}$ the support of ξ^* and $\mathcal{S}^*$ a subset of $\mathcal{S}_{\xi^*}$. Denote $\eta'(x,\beta) = \eta[x,\theta(\beta)]$. Since $\partial\eta'(x,\beta)/\partial\beta = \mathbf{J}^\top\partial\eta(x,\theta)/\partial\theta$ and $\mathbf{J}^\top\mathbf{c} = (1, \mathbf{0}_{p-1}^\top)^\top$, we obtain

$$\int_{\mathcal{S}^*} \frac{\partial\eta'(x,\beta)}{\partial\beta} \, \xi^*(dx) - \int_{\mathcal{S}_{\xi^*}\setminus\mathcal{S}^*} \frac{\partial\eta'(x,\beta)}{\partial\beta} \, \xi^*(dx) = \gamma[\theta(\beta)] \begin{pmatrix} 1 \\ \mathbf{0}_{p-1} \end{pmatrix} .$$

Therefore, $\int_{\mathcal{S}^*} \eta'(x,\beta)\,\xi^*(dx) - \int_{\mathcal{S}_{\xi^*}\setminus\mathcal{S}^*} \eta'(x,\beta)\,\xi^*(dx) = G(\beta_1)$, with $G(\beta_1)$ some function of β_1, estimable for ξ^*. Finally, $\beta_1 = \mathbf{c}^\top\theta$ is locally estimable for ξ^* since $G(\beta_1)/d\beta_1 = \gamma[\theta(\beta)] > 0$. ■

Example (continued): Take $\mathbf{c} = (0,\ 1)^\top$, $\mathbf{c}^\top\theta$ is not locally estimable at $\theta^0 = (0.7,\ 0.2)^\top$ for the c-optimal design $\xi^* = \delta_{x_*}$, with $x_*(\theta^0) \simeq 4.28$, but the value of x_* depends on θ^0 through (20). Taking two different nominal values $\theta^{0,1}, \theta^{0,2}$ is enough to construct a nonsingular design by mixing the associated c-optimal designs. □

References

[1] H.J. Bierens. *Topics in Advanced Econometrics.* Cambridge University Press, Cambridge, 1994.

[2] G. Elfving. Optimum allocation in linear regression. *Annals Math. Statist.*, 23:255–262, 1952.

[3] R.I. Jennrich. Asymptotic properties of nonlinear least squares estimation. *Annals of Math. Stat.*, 40:633–643, 1969.

[4] J. Kiefer. Two more criteria equivalent to D-optimality of designs. *Annals of Math. Stat.*, 33(2):792–796, 1962.

[5] E.L. Lehmann and G. Casella. *Theory of Point Estimation.* Springer, Heidelberg, 1998.

[6] A. Pázman. *Foundations of Optimum Experimental Design.* Reidel (Kluwer group), Dordrecht (co-pub. VEDA, Bratislava), 1986.

[7] A. Pázman and L. Pronzato. Asymptotic criteria for designs in nonlinear regression. *Mathematica Slovaca*, 56(5):543–553, 2006.

[8] A. Pázman and L. Pronzato. On the irregular behavior of LS estimators for asymptotically singular designs. *Statistics & Probability Letters*, 76:1089–1096, 2006.

[9] A. Pázman and L. Pronzato. Asymptotic normality of nonlinear least squares under singular experimental designs. In L. Pronzato and A.A. Zhigljavsky, editors, *Optimal Design and Related Areas in Optimization and Statistics.* Springer, 2008. to appear.

[10] C.F.J. Wu. Asymptotic theory of nonlinear least squares estimation. *Annals of Statistics*, 9(3):501–513, 1981.